UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK

BARBARA SCHWAB et al., individually and on behalf of all others similarly situated, Plaintiffs, v. PHILIP MORRIS USA, INC. et al., Defendants.	Case No. CV-04-1945 (JBW) (SMG)

Draft Expert Witness Report by:

Dr. John R. Hauser
MIT Sloan School of Management
Massachusetts Institute of Technology, E56-314
Cambridge, Massachusetts 02142
(617) 253-2929 Fax: (617) 258-7597
jhauser@mit.edu; web.mit.edu/hauser/www; mitsloan.mit.edu/vc

Table of Contents

I. Introduction and Qualifications

1. My name is Dr. John R. Hauser. I am the Kirin Professor of Marketing at the MIT Sloan School of Management at the Massachusetts Institute of Technology ("MIT"). I have served MIT in a number of capacities including Head of the Marketing Group, Director of the Center for Innovation in Product Development, and Director of the International Center for Research on the Management of Technology. I have recently been appointed Area Head for Management Science at MIT. The Management Science Area at the MIT Sloan School of Management includes the Marketing Group, the Statistics Group, and other groups. The principal focus of my research and teaching at MIT has been in the areas of marketing management, new product and service development, consumer satisfaction, marketing research, and competitive marketing strategy.
2. I am the author of over sixty articles and papers, as well as the textbooks *Design and Marketing of New Products* and *Essentials of New Product Management*. In addition, I served as editor-in-chief of *Marketing Science* and have held senior editorial positions with *Management Science* and the *Journal of Product Innovation Management*. I have also received numerous awards for excellence in research and teaching in marketing and marketing research, and was recognized by the American Marketing Association with the Converse Award for "outstanding contributions to the development of the science of marketing." In September of 2001 I received the Parlin Award, "the oldest and most distinguished award in the marketing research field," according to the American Marketing Association.[1] I am a trustee of the Marketing Science Institute.
3. I have served as an expert witness in connection with a range of disputes. Most of this

[1] See www.marketingpower.com and type in Parlin Award in search box. Click the link that says "Parlin Award." The direct link is http://www.marketingpower.com/live/content.php?Item_ID=1097. Visited July 7, 2005.

expert testimony has involved surveys and other market research to measure consumers' attitudes, beliefs, and intentions. I have been called upon to project what consumers would have done in different market scenarios, to measure the importance of product features, to measure the impact of rumors, to evaluate marketing research with respect to advertising claims, and to investigate the potential for consumer confusion. I have also consulted to dozens of major corporations, including General Motors, Fidelity Investments, American Airlines, Procter & Gamble, and IBM. My professional qualifications are described in my curriculum vita, which is attached as Exhibit A. A list of cases in which I have testified within the last four years at deposition or trial is attached as Exhibit B.

II. Assignment

4. I was asked by counsel for Plaintiffs to assess the value and importance of health risks to "light" cigarette consumers in their decision to purchase a "light" cigarette. I assess this value and importance in the first study, which I label the Conjoint Study.
5. I was also asked to attempt to design and implement a survey research study to provide evidence with which to assess "how many smokers of "light" cigarettes were ignorant of the alleged fraud in each relevant year." If, based on initial tests, smokers could provide reliable and valid responses to retrospective questions, I would then undertake a more comprehensive study.
6. In undertaking this assignment, I relied on my extensive expertise in developing, testing, and analyzing surveys and in interpreting qualitative and quantitative research about consumer attitudes, intentions, and behavior.

7. My work is ongoing; I may update and revise my results and conclusions as I review additional data and information. A complete list of materials I have considered to date in connection with this particular assignment is included as Exhibit C. To the extent that I review additional information, I will supplement this list.
8. Part of the work for this investigation was performed under my direction by others at Applied Marketing Science, Inc. ("AMS"). I am a Senior Consultant for and Co-Founder of AMS.
9. My rate of compensation for this task is $650 per hour. My compensation is not contingent upon the outcome of this dispute.

III. Summary of Conclusions

10. The scientific methodology used to design, execute, and analyze the Conjoint Study in this report is sound, reliable, and valid. The results can be relied upon to draw inferences about whether health risks are a significant contributing factor in consumer decisions to smoke "light" cigarettes and what proportion of "light" cigarette-smoking consumers relied on health risks as a significant contributing factor. The results can further be relied upon to draw inferences about how consumers and the market would react to cigarettes with different levels of health risks.
11. Based on the methodologies described in this report and based on calculations of the importance of health risks as a factor in consumer decisions, consumer willingness to pay for lower health risk, and market-based simulations, I conclude that health risks are a positive contributing factor in the choice of "light" cigarettes for 90.1 percent of "light" cigarette consumers and, of these consumers, on average, it is ranked above all other measured features, excluding price, (i.e., taste or packaging). It is ranked above at least one other feature by 97.8 per-

cent of the consumers, indicating its significance for the overwhelming majority of respondents.[2]

12. The methodologies described in this report provide data with which to calculate the market value of the change in perceived health risks for "light" cigarettes. In this report I provide two examples of how one might calculate the market value. These examples suggest that the market value of a decrease in health risks from the same as regular cigarettes to the same as "light" cigarettes is between 39.8 percent and 47.3 percent of the price per pack.

13. For "light" cigarette consumers, the data and analyses in this report can be used to simulate "but for" scenarios in which "light" cigarettes with varying health risks are available on the market. For example, if each brand of "light" cigarettes offered two versions of "light" cigarettes, one version with the perceived health risks the same as "light" cigarettes at regular price and another version with the perceived health risks the same as regular cigarettes but at a 50% reduction in price, then the "light" cigarettes with the better perceived health risks would obtain a market share between 46 and 48 percent.[3]

14. The data and analyses in this report can be used to gain insight on the price discount that would be needed to sell a cigarette with perceived health risks greater than regular cigarettes. For example, more than 75 percent of the consumers would be willing to pay more than 50 percent of the price per pack to decrease health risks from greater than regular cigarettes to health risks the same as "light" cigarettes. One estimate of the market value of a perceived decrease in health risks from greater than that of regular cigarettes to that the same as "light"

[2] Health risks are ranked above price by 26.8 percent of these consumers, above taste by 68.9 percent of the consumers, and above pack type by 94.4 percent of the consumers. Health risks are ranked above price or taste by 76.1 percent of these consumers.

[3] The estimates are 47.7 percent for a market simulation based on first choices and 46.1 percent for a market simulation based on randomized first choices.

cigarettes is substantially more than 50 percent of the price per pack.[4]

15. At most 8/10^{ths} of 1 percent of the respondents use a non-compensatory lexicographic decision rule for taste, health risks, pack type, and price. For all other respondents and for the features of taste, health risks, and price, high levels on some features can compensate for low levels on other features.

16. I followed accepted scientific methodology to design, execute, and analyze the Experimental Time Study. In addition, I focused on those respondents who, a priori, I believed to be more able to provide estimates as to the times at which they changed their beliefs about health risks and the times at which they first came to believe that they had been misled by the cigarette companies about health risks.

17. The results of the Experimental Time Study suggest that it is difficult for respondents to estimate the year in which they changed their beliefs about health risks and the year in which they first came to believe that they had been misled by the cigarette companies. I do not have confidence in the estimates of these years as provided by respondents.

IV. Overview of Methodology – Conjoint Study

18. The basic methodology that I selected is known as web-based conjoint analysis. Conjoint analysis is a tool that is widely used in the field of marketing research. It was introduced to the field of marketing research in 1971 and is generally recognized by marketing science academics and industry practitioners to be the most widely studied and applied form of quan-

[4] The largest price discount examined in the survey was a 50% percent discount, hence, conservatively, I do not simulate any market with a price in which a brand uses a discount of more than 50%. If each brand of "light" cigarettes offered two versions of "light" cigarettes, one version with perceived health risks the same as "light" cigarettes at regular price and another version with perceived health risks greater than regular cigarettes but at a 50% reduction in price, then the "light" cigarette with better perceived health risks would obtain a market share between 72 and 76 percent. The estimates are 75.4 percent for a market simulation based on first choices and 72.3 percent for a market simulation based on randomized first choices. Details are provided later in this report.

titative consumer preference measurement. It has been shown to provide valid and reliable measures of consumer preferences, and these preferences have been shown to provide valid and reliable forecasts of what consumers will do (or would have done) under scenarios related to those measured.[5] For example, under the auspices of MIT's Virtual Consumer Initiative, my colleagues and I have undertaken large-scale tests of the validity of web-based conjoint analysis. Predictions were highly accurate. One of the scientific papers discussing the validity test recently received two highly prestigious awards as the best paper in the marketing sciences literature for 2003 (awarded in 2004) and for the best paper based on a dissertation (awarded in June 2005).[6] Another scientific paper was a finalist for the best paper in 2002 in the *Journal of Product Innovation Management* and still a third paper was a finalist for the best contribution to the practice of marketing research in 2004.[7]

19. The general idea behind conjoint analysis is that consumers' preferences for a particular product are driven by features or descriptions of features embodied in that product. For example, a cigarette might be described by features such as: (i) pack type (hard or soft); (ii) degree of perceived health risks; (iii) taste; and (iv) price. A feature, such as perceived health risks, can have many levels such as: less than "ultra-light" cigarettes, the same as "ultra-light" cigarettes, the same as "light" cigarettes, the same as regular cigarettes, and greater than regular cigarettes. (Detailed wording and descriptions of these levels are provided later in this report.) When applying (decompositional) conjoint analysis, respondents are asked to make holistic judgments about products (or product descriptions) or to choose among prod-

[5] Hauser, John R. and Vithala Rao (2004), "Conjoint Analysis, Related Modeling, and Applications," *Advances in Marketing Research: Progress and Prospects*, Jerry Wind and Paul Green, Eds., (Boston, MA: Kluwer Academic Publishers).

[6] Toubia, Olivier, Duncan I. Simester, John R. Hauser, and Ely Dahan (2003), "Fast Polyhedral Adaptive Conjoint Estimation," *Marketing Science*, 22, 3, (Summer), 273-303.

[7] Dahan, Ely and John R. Hauser (2002), "The Virtual Consumer," *Journal of Product Innovation Management*, 19, 5, (September), 332-354;Toubia, Olivier, John R. Hauser, and Duncan Simester (2004), "Polyhedral Methods for Adaptive Choice-based Conjoint Analysis," *Journal of Marketing Research*, 41, 1, (February), 116-131.

ucts (or product descriptions). That is, consumers are shown product profiles made up of the features or descriptions of features and asked to indicate their preferences for these profiles or to choose among these profiles ("choice task"). The conjoint analysis methods use the holistic judgments or choice tasks to decompose respondent preferences for a product into the partial contribution of these feature levels or descriptions ("partworths"). These partworths are then estimated from respondent preferences or choices with the appropriate statistical methods. The partworths for feature levels are identified with the estimation methods so that the partworths best predict consumer preferences or choices. The difference between the smallest and largest partworths for levels of each feature can be used to calculate the relative importance of each feature in purchase decisions. Importance refers to the relative value of changing that feature from its least preferred level to its most preferred level. The partworths for changes in price levels measure the relative importance of changing price. The price reduction needed to compensate for a lower (less-preferred) level of a feature, or the additional price consumers would pay for the higher (more-preferred) level of feature can then be calculated.

20. There are many forms of conjoint analysis, most of which provide valid and reliable data. For this assignment, I selected a form of conjoint analysis known as Choice-Based Conjoint ("CBC") analysis. In CBC, consumers are shown sets of profiles (called the "choice sets"), and asked simply to choose the profile that they most prefer, that is, the profile that they would choose if the choice set described the only products that were available. I chose to show respondents four products in each choice set. I have used four-product choice sets in other applications and have found the data to be both reliable and valid.

21. Choice-Based Conjoint analysis is consistent with economic theories of approximate utility

maximization. That is, if a researcher could measure each and every feature of the product and represent consumer utility as a function of those features, then consumers would choose the product that maximizes their utility. This utility is composed of the features that are measured (as represented by the partworths) and features that are not measured (random component). To estimate the partworths, I use a statistical method known as Hierarchical Bayes ("HB"). HB is based on "Bayesian" methods. In lay terms, a Bayesian method uses the data, that is, the respondent's choices in the questionnaires, to update any prior beliefs, such that the resulting partworth estimates make the best use of the data. The "hierarchy" part of HB means that the estimates for a given respondent are based on the choices by that respondent and informed by the choices of other respondents. This information is used iteratively, so that the resulting partworth estimates most accurately reflect all of the data in the sample. HB has proven to provide reliable and valid conjoint analysis estimates of partworths. It is an appropriate method to use to obtain partworths when there are a moderate number of choice sets in the choice task. This enables me to appropriately balance the number of questions in the choice task with the number of partworths that need to be estimated. I have observed in other conjoint surveys that I have conducted that such designs limit respondent wear-out.[8]

22. The form of partworth estimation that I selected allows consumer preferences to be heterogeneous. That is, each respondent in the sample can have different values for his or her partworths. For example, one respondent might prefer the taste of regular cigarettes and another might prefer the taste of "light" cigarettes. One respondent might value having lower health risks highly and another less highly. In technical terms, I estimate a separate "vector" of

[8] If too many questions were asked of a respondent, then the respondent might "wear out," that is, response errors might increase as the respondent tires. Not only did I limit the number of questions in the choice task to minimize wear out, but I pretested the questionnaire to ensure that respondents did not experience wear out.

partworths for every respondent.[9]

23. Based on the measured features and random component, I predict the probability that a respondent will choose any product profile that is described by the partworths and can do so for any competitive set of products described by the measured features. The probability-of-choice calculations are based on an "extreme-value" distribution, and the analysis method is commonly referred to as the multinomial logit model. This is the appropriate distribution to use in this context because, in making their selections, consumers are maximizing their utility. When price is one of the measured features, the value consumers place on each of the other features can be expressed in terms of price. That is, the price reduction needed to compensate for having a lower level of a feature, or the additional price consumers would pay for having a higher level of a feature can be calculated. I also simulate how choice shares would change in a market based on a change in overall price.

24. Because CBC is based directly on consumer choices, it is, in my opinion, an ideal method to determine the value that consumers place on the various features of cigarettes. In particular, CBC can determine the value that consumers place on having reduced health risks as part of their "light" cigarette. CBC can also assess the significance of health risks as a contributing factor in consumers' decisions to smoke "light" cigarettes.

V. Questionnaire Development – Conjoint Study

25. I began by identifying the features that drive "light" cigarette consumers' purchases of cigarettes. I instructed AMS to conduct in-depth interviews with current "light" cigarette consumers. A total of 14 interviews were conducted on March 9th, 2005. From these interviews,

[9] Technically, we also have a full characterization of the (posterior) distribution of partworths for each respondent. This full characterization is used in the randomized first choice method described later in this report.

both AMS and I learned more about the features of cigarettes and how these features affect consumers' purchasing decisions. These interviews enabled me to identify the appropriate features to use in the web-based, Choice-Based Conjoint analysis and to develop a questionnaire that used words and phrases that consumers use to describe the features of cigarettes. These interviews ensured that the words used to describe the levels of the features were understood by consumers. This questionnaire was programmed into a web-based software system designed for administering and analyzing such questionnaires.[10] Examples of the final questionnaires that respondents were asked to complete are shown in Exhibit D. Recall that respondents answered these questions via their computers. Exhibit D contains reproductions of the computer screens. In addition, some questions, such as the choice task, were chosen based on algorithms that included appropriate randomization to avoid order effects. Thus, Exhibit D is an example of the types of screens that respondents viewed.

VI. Pretesting the Questionnaires – Conjoint Study

26. The questionnaires were pretested with 9 respondents on March 25th through March 29th, 2005 to ensure that respondents understood the descriptions, instructions, and questions and that their answers adequately represented their beliefs. Minor changes in the wording and formatting of the questions were made as a result of the pretest to ensure that respondents understood the questions and that the interview flowed smoothly. Respondents were debriefed to ensure that the questionnaire maintained a "double-blind" protocol. In a double-blind protocol, neither the interviewer nor the respondent is given either explicit or implicit cues from which to guess the purpose of the study. In this case, the "interviewer" is the web-

[10] I used Sawtooth Software, Inc.'s SSI Web Version 3.5.0 package, which is a well-known and widely used software system for these types of applications.

based questionnaire, so we need debrief only the respondents. Following standard procedures, no pretest responses were included in the final sample.

VII. Identifying the Sample – Conjoint Study

27. For this survey, potential respondents were selected at random from Greenfield Online's database and sent an invitation (Exhibit E) to go to a special website to complete the survey. Each invitation included a URL with an embedded password that was then matched against a list of valid passwords and against the list of passwords that had already been used. (The former ensures that only valid respondents complete the questionnaire. The latter ensures that each respondent completes the questionnaire at most once.) Respondents received an initial e-mail invitation and up to four e-mail reminders. Greenfield Online motivates respondents to participate in these surveys by adding $5 to the Greenfield prize accounts of all who qualify for and complete the survey. In my experience, such incentives increase response rates but do not bias any of the responses to the questions in the survey.

28. In order to qualify for the survey, respondents were screened to ensure that they were "light" cigarette consumers. A total of 627 respondents completed the survey beginning on June 15th, 2005 and ending on June 29th, 2005. The completion rate was 94.9 percent. Details are provided in Exhibit F.

29. To ensure a nationally representative sample of respondents, quotas were set so that the sample would match the national data on Census region, sex, age, and household income. These respondents were allocated to the seventy-two quota groups (four Census Regions crossed with two levels of sex, three levels of age, and three levels of household income, 72 =

4x2x3x3).[11] Based on information from Greenfield Online, I estimated that 12.5% of the respondents would be "light" cigarette consumers. In order to identify approximately 500 "light" cigarette consumers to interview, the study would need to identify approximately 4,000 respondents allocated proportionally to the quota groups. In order to identify these 4,000 potential respondents, 7,738 potential respondents were screened. Exhibit F shows the number of respondents in each group and the number completing the conjoint task. The completed interviews match closely the screening quota and the census categories.[12]

30. In my experience one obtains the same partworth estimates (up to normal sampling variation) from web-based and from central facility respondents. For this assignment, I followed protocols designed to maximize the response rates to the surveys.[13] It is my opinion that these protocols are sufficient to ensure that the respondents are representative of the sampled population.

VIII. Survey Administration – Conjoint Study

31. After the initial screening for representativeness, respondents were asked whether they smoke cigarettes and, if so, which type they primarily smoke. Only respondents who were "light" cigarette consumers were asked to continue the interview.

32. Respondents were asked how long they had been smoking, what type of cigarette they had smoked primarily when they began smoking, and how long ago they began to smoke primarily "light" cigarettes. They were asked how many packs per day they smoke and asked to in-

[11] Because the respondents are smokers, the respondents must be 18 years of age to participate.
[12] There is no reason to expect that the penetration of "light" cigarette consumers will match the distribution of census regions and/or the distribution of age, sex and household income. The screening ensures we start out with a representative sample. The final sample reflects the distribution of "light" cigarette consumers as observed.
[13] A complete disposition of the samples for the survey is provided in Exhibit F. Based on these calculations, the incidence rate for the survey was 17.4 percent, the initial response rate was 15.6 percent, the completion rate for the conjoint task was 94.9 percent, and the net response rate was 14.8 percent.

dicate which brand of "light" cigarette is their primary brand. Respondents were then introduced to the conjoint task and shown a list of the features that would be varied in the product profiles that were to follow.

33. Respondents were then shown a series of sixteen screens (choice sets) containing four alternative cigarette options that were described by the combinations of the features that had been identified by qualitative research. For each set of four alternative cigarette options, respondents were asked, "If these were your only options, which would you choose?" Prior to the choice exercise, respondents were instructed to choose from each choice set the cigarette option they would choose as their primary cigarette option. Respondents indicated which of the four cigarette options they would choose. The features and feature levels are indicated below. Recall that these features and levels were chosen to be realistic based on both the qualitative interviews and the pretest interviews.

    - Pack type. The levels were soft pack and hard pack.[14]
    - Level of perceived health risks. The survey emphasized that consumers were to use their personal beliefs about health risks in the choice exercise.[15] The levels were "Health risks are greater than regular cigarettes," "Health risks are the same as regular cigarettes," "Health risks are the same as "light" cigarettes," "Health risks are the same as "ultra-light" cigarettes," and "Health risks are less than "ultra-light" cigarettes."
    - Taste. The levels were "Tastes like a regular cigarette," "Tastes like your brand of "light" cigarette," and "Tastes like an "ultra-light" cigarette." Respondents were

[14] Qualitative interviews and pretests suggested that the inclusion of a feature such as pack type enhanced the realism of the survey.

[15] The survey uses two-sided introductory statements to emphasize that the respondent is to consider his or her personal beliefs. Through qualitative interviews and pretests, I found that these descriptions were well-understood by respondents and that they understood they were to use their perceived beliefs about health risks.

asked to use their own beliefs about taste, and to assume the taste for a cigarette option was the same as the taste provided by their brand for that cigarette option (to the degree possible).[16]

- Price per pack. The levels were "50% less than what you pay now", "20% less than what you pay now", "The same price that you pay now", "20% more than what you pay now" and "50% more than what you pay now" per pack.[17]

34. For each respondent, his/her choices from the choice sets he/she was shown were used to estimate with hierarchical Bayes (HB) the vector of partworths for each respondent.[18] The options in the choice sets were chosen randomly to ensure equal expected values of each feature and level. The designs were highly efficient.[19] Intuitively, in our study, an efficient design provides the estimates of partworths with high precision. Efficiency refers only to precision.

35. Following these sixteen choice sets, respondents were thanked and the interview was concluded.

IX. Analysis – Conjoint Study

36. Hierarchical Bayes estimates for the partworths for each respondent were obtained from the

[16] Not all brands of "light" cigarettes offer regular and ultra-light versions. Thus, I did not use the words "your brand of" for regular and ultra-"light" cigarettes. Rather, I used a paragraph description in which the survey told the respondent that, if the respondent's brand of "light" cigarette offers a regular (ultra-light) version, then the respondent should consider taste like that version. Otherwise, I listed examples of popular cigarettes. Through qualitative interviews and pretests, I found that these descriptions were well-understood by respondents.

[17] To aid respondents, the survey provided examples of how to convert percentage reductions (increases) to dollar amounts. Through qualitative interviews and pretest, I found that these descriptions and calculations were clear and well-understood by respondents.

[18] Technically, this is a Bayesian procedure in which we are obtaining the best update of the partworths based on the data. This means that we represent all of the information about the partworths, including the mean (average) for each respondent and a measure of the uncertainty about the partworths. For brevity I refer to these complicated descriptions as simply the estimates of the partworths.

[19] For more technical descriptions see Sawtooth Software Technical Paper, "Choice-based Conjoint (CBC) Technical Paper," 2001. Efficiencies were 100% for the complete-enumeration randomized designs.

data with software developed by Sawtooth Software, Inc.[20] The software uses a Bayesian procedure to update (estimate) the relative values of the partworths for each respondent. These heterogeneous partworths are used in the simulation model. In Exhibit G, I summarize the (average) partworths where the average is computed over all respondents within a group.[21] Exhibit G also shows the standard deviation of the estimated partworths (across respondents), which demonstrates the heterogeneity in respondents. (The higher the standard deviation, the more the partworths vary over respondents. It is natural that partworths vary among respondents. This indicates that different respondents may value the various features differently. As described in subsequent paragraphs of this report, I take into consideration this heterogeneity in my analyses.) When interpreting Exhibit G it is also useful to keep in mind that the partworths do not vary independently across respondents. Technically, there are also covariances (across respondents) among the various partworths. For example, respondents who place a higher importance on perceived health risks also place a lower importance on taste, as indicated by a -0.22 correlation between the full-range importances of these features.

37. The average values of the estimated partworths in Exhibit G indicate that, on average, "light" cigarette consumers value lowering their health risks. That is, the average of the partworth for having lower health risks than an "ultra-light" cigarette is statistically significantly larger than the average partworth for having health risks greater than a regular cigarette.[22] This is also true for health risks the same as "light" cigarettes compared to health risks to the same

[20] Sawtooth Software CBC/HB Module for Hierarchical Bayes Estimation, Version 3.1.
[21] In conjoint analysis we are dealing with choices among alternative products. Thus, partworths indicate the value to a respondent in changing one level of a feature for another. For example, our analyses measure the value of having a soft pack relative to having a hard pack. For example, based on the averages in Exhibit G this relative value is 12.7, that is, 6.35 – (-6.35) = 12.7. Because the relative values are used in the forecasting model, we report the mean partworths such that they sum to 0.00 across feature levels within a feature. For example, 6.35 + (-6.35) = 0.00.
[22] t-statistic = 33.9 and $p \leq 0.00$

as regular cigarettes.[23] The values of the partworths for health risks do not vary significantly among users of the major brands of "light" cigarettes.[24]

38. In order to establish the appropriateness of using the estimated partworths to forecast consumer behavior, I test the fit and predictive ability of the conjoint analysis estimates. An appropriate statistic with which to evaluate the model is the percentage of uncertainty that is explained by the model.[25] This statistic is known as "U^2." U^2 is a ratio of the "information explained by the probabilistic model" divided by "the total uncertainty (entropy) of the system." Perfect predictions would have a U^2 of 1.0. A "null" model, which predicts that each respondent would choose randomly among the four profiles in each choice set (predicted probability = 0.25 for every profile), would have a U^2 of 0.0. I examine U^2 for the choice tasks that were used to estimate the model and for holdout choice tasks (choice tasks that were "held out" of the estimation in order to test predictive ability).

39. To get a valid indicator of holdout performance, I used the HB method excluding one choice task for each respondent from the estimation. I repeated this process three times using a different choice task each time. The resulting statistics are the average of the three separate validation calculations. If U^2 for the estimation profiles is at a reasonable level, then it is appropriate to conclude that the HB estimates explain a reasonable level of consumer behavior and that the features as specified capture much of consumer preferences. (U^2 measures the ability of the model to explain consumer choices. The partworths and the forecasts are still the best estimates given the data.) To establish "reasonable" I compare the HB estimates to those obtained by a multinomial logit model in which the partworths are the same for every

[23] t-statistic = 31.9 and $p \leq 0.00$
[24] F-statistic = 1.094, p = .222, using Wilk's Lambda method.
[25] Hauser, John R. (1978). "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach, *Operations Research*, Vol.26, No. 3 (May-June), 406 – 421.

respondent (homogeneous multinomial logit or "HML"). I also examine whether or not the HB estimates "over-fit" the data. "Over-fitting" would occur if the estimates from the estimation partworths could not predict the choices in the holdout data to a reasonable level of precision.

40. For the HB estimates obtained from estimation on 15 of the 16 choice tasks, the average estimation U^2 is 0.522.[26] The average U^2 for the holdout task is 0.459. For the model in which the partworths are the same for every respondent (HML), the estimation U^2 is 0.288 and the holdout U^2 is 0.279. Thus, the HB estimates are substantially better than the HML estimates. This is true for both the estimation data and the holdout data. Furthermore, there is only a modest drop off in predictive ability for the holdout data suggesting that the HB estimates are not over-fitting the data. Finally, the HB-based U^2 value for the holdout task is substantially above the HML-based U^2 value and substantially above that which would be obtained from random choice. Based on these statistics, it is my opinion that the HB estimates are appropriate for making predictions with respect to alternative scenarios.

41. Because U^2 is a technical statistic, it is informative to look at a more intuitive measure, that is, the percentage of choices that can be predicted correctly with the HB estimates. For this statistic, the "null" model of random choice would predict the choice correctly only 25 percent of the time (one time out of four). In this case, the estimated partworths correctly predict the chosen alternative 87.6 percent of the time for the estimation data and 72.3 percent of the

[26] In calculating U^2 for the estimation sample it is appropriate to adjust for the degrees of freedom in the model. This is appropriately conservative because it lowers the U^2 value. In an HB model we need to estimate the degrees of freedom as affected by the hierarchical constraints. This is a technical calculation involving the Deviation Information Criterion (DIC). We adjust the degrees of freedom by lowering the likelihood value by the degrees of freedom and then computing a ρ^2 fit statistic. See Bruce G. Hardie, Eric J. Johnson, and Peter S. Fader (1993), "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 12, 4, (Fall), 389. Because Hauser (1978) demonstrates that ρ^2 is numerically equal to U^2 this correction for degrees of freedom can also be used for U^2 while retaining the intuitive interpretation of U^2.

time for the holdout data.[27] By contrast, the homogeneous multinomial logit (HML) analysis correctly predicts the choice only 59.5 percent of the time for the estimation data and 60.7 percent of the time for the holdout data. For both the estimated data and for the holdout data, the HB estimates are significantly better at predicting choice than are the HML estimates at the 0.00 level. Furthermore, both the HB and the HML estimates are significantly better at predicting choice than the null model at the 0.00 level.[28]

42. The statistical tests based on the statistic of "percentage of choices correctly predicted" lead to the same conclusions as the more formal U^2 tests. It is my opinion that the HB estimates are appropriate for making predictions with respect to alternative scenarios.

43. In an HB CBC analysis of consumers, I estimate the partworths that describe the process by which consumers choose among alternative cigarette options. Technically, this is an additive model because we sum the partworths of the feature levels in order to estimate the utility of an option.[29] An additive model is a general representation. The partworths can represent both "compensatory" and "non-compensatory" decision rules. In a compensatory decision rule, improvements in some features can compensate for decreases in other features. For example, in a compensatory rule, we might find that better taste, better health risks, and lower price might compensate for a less-preferred pack type. In a non-compensatory rule, improvements in some features will not compensate for a lowered level of a more important feature. A common type of non-compensatory rule is a lexicographic rule in which a con-

[27] Unlike the more rigorous U^2 statistic, there is no easy way to adjust the first-choice prediction percentage for degrees of freedom. Thus, we consider the more conservative holdout percentage of 72.3%.

[28] When comparing HB and HML, t-statistic = 96.0 for fit and 8.70 for holdout; when comparing HB and the null model, t-statistic = 192.4 for fit and 32.9 for holdout; when comparing HML and the null model, t-statistic = 87.4 for fit and 23.24 for holdout. (The null model is a vector of hits with a 25 percent hit rate.) For all of these t-statistics, $p \leq 0.00$. Data for the fit t-tests were based on 627 respondents x 15 choice tasks x 3 different holdout tasks (Choice tasks 4, 8, or 16) = 28,215 observations. The hits data, a series of ones and zeros, were formed into 28,215 x 1 column vectors for HB and HML. For the holdout cells, there were 627 respondents x 1 choice task x 3 different holdout tasks = 1881 observations. Paired t-tests were performed on the resulting vectors for HB vs. HML.

[29] I also tested a model with full interactions. The additive model is a better description of consumers for these data.

sumer first chooses based on one feature, say pack type, then another, say taste, then another, etc. until only one option remains. A lexicographic rule is similar to alphabetical order. The word "azygous" comes before "babble" in the dictionary because "a" is before "b" in alphabetical order. The "a" and two additional "b's" in babble do not compensate even though azygous has a "z," a "y," and a "u." Technically, an additive model represents a compensatory rule when the relative partworth values are not too extreme. An additive model can represent a lexicographic rule when the partworths are extreme. I address what I mean by "extreme" in subsequent paragraphs.

44. Because an additive model represents both compensatory and non-compensatory rules, the CBC analyses in this report are appropriate whether or not respondents are compensatory. However, I can use the estimated partworths to examine whether or not improvements to some features of cigarettes can compensate for less-preferred levels of other features. The use of compensatory, non-compensatory, and lexicographic rules is an empirical question that depends on the product category. For example, in a recently completed study, I found that for the choice of SmartPhones, many respondents use lexicographic rules such as rejecting an extremely high price.[30]

45. If a consumer is using a lexicographic rule, then no improvements in lesser important features can compensate for any improvement in the most important feature. Suppose that pack type is the most important feature and suppose the consumer prefers a hard pack. Then the consumer is lexicographic with respect to pack type if the importance of pack type (the partworth of a hard pack minus the partworth of a soft pack) is greater than the sum of the importances of taste, health risks, and price. For features with multiple levels, the formulae are

[30] Yee, Michael, John Hauser, James Orlin, and Ely Dahan (July 2005), "Greedoid-Based Non-compensatory Two-Stage Consideration-then-Choice Inference," under review, *Marketing Science*.

complicated but can be computed.[31]

46. Using the partworths estimated by HB CBC, I examined whether consumers are lexicographic with respect to pack size, taste, health risks, or price. I first identified the most important feature for each respondent and then compared the partworth difference for any change in that feature to the sum of the importances for the other features. By this test, 622 of the 627 respondents were not lexicographic (99.2%). The remaining five respondents exhibited lexicography with respect to pack type.[32] No one was lexicographic (at the top level) with respect to taste, health risks, or price. In other words, for 99.2 percent of the respondents and for the features of taste, health risks, and price, improvements in some features can compensate for decreases in other features.

47. Using the HB partworths, I can draw conclusions about whether perceived health risks are a significant contributing factor in "light" cigarette consumers' purchase decisions.[33] I begin by examining the importances that consumers place on health risks. In the HB CBC analysis, the importance of health risks is the difference between the partworth for "Health risks are less than an "ultra-light" cigarette" and "Health risks are greater than a regular cigarette." For 90.1 percent of the 627 respondents, this difference was positive, indicating that, given the choices made by respondents who answered the survey, the best estimate of their importance

[31] For detailed formulae see Yee, et. al. (ibid) and Jedidi, Kamel and Rajeev Kohli (2004), "Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers," Working Paper, Graduate School of Business, Columbia University (November).

[32] For these five respondents, after choosing their favorite pack type lexicographically, the remaining three features were evaluated with a compensatory decision rule.

[33] When doing so, it is important to distinguish between the legal term "significant" and the statistical term "significant." Because I am not a lawyer, I provide no legal opinion with respect to the legal term "significant." I provide expert opinions about percentages. When I use the term "significant" in a statistical sense, I refer to the probability that the observed outcome could be due to chance alone. It is common in scientific studies to set minimum level of significant level of 0.05, which means that there is a 5 percent chance or less that the observed data are due to chance. (Some scientists may also require a 0.01 level and others may report results at the 0.10 level with the appropriate caveats. Many of the statistical significance levels in this report are significant at better than the 0.01 level as indicated.) When I use the term "significant" in a non-statistical sense I am relying on counsel to provide a legal interpretation.

for health risks is positive. I study this more closely by examining statistical significance. In particular, for some of the remaining 9.9% of the respondents the choices might be such that we cannot be statistically confident that the importance they place on health risks is different than zero. Because we are testing whether the importance is positive (or, in other tests, negative), the appropriate statistical test is a one-tailed statistical test.

48. The HB analysis provides both means and standard deviations of the partworths for every respondent. I use these means and standard deviations to perform the appropriate one-tail t-tests (statistical test) for each respondent in the sample. For 76.4 percent of the respondents, the importance of health risks is positive and statistically significant at the 0.10 level and for 69.7 percent of the respondents the importance of health risks is positive and statistically significant at the 0.05 level. For only 2.1 percent of respondents the importance of health risks is negative and statistically significant at the 0.10 level and for 1.1 percent of the respondents the importance of health risks is negative and significantly negative at the 0.05 level. For the remaining respondents (29.2 percent at the 0.05 level, 21.5 percent at the 0.10 level), at the level of the individual respondent, the data are not sufficient to classify respondents at a high level of statistical significance. In summary, at the 0.05 level, I am statistically confident that about 1 percent of the respondents do not place a significant positive value on health risks. For almost 70 percent of the respondents I am statistically confident that they place a significant positive value on health risks. For the remaining respondents, approximately 20%, my best estimate is that about 2/3rds of the respondents place a positive value on health risks – for a total of 90.1% as summarized in the previous paragraph.

49. I further address whether perceived health risks are a significant contributing factor in "light" cigarette consumers' purchase decisions by ranking importance measures of the feature. (In

this paragraph I use the word "significant" in a non-statistical sense.) I first consider only those respondents for whom I estimate that the respondent places a positive value on health risks. I next compute the importance of each feature as the difference between the maximum partworth for that feature and the minimum partworth for that feature. I then rank these feature importances. Health risks are ranked above price by 26.8 percent of these consumers, above taste by 68.9 percent of the consumers, and above pack type by 94.4 percent of the consumers. Health risks are ranked above price or taste by 76.1 percent of these consumers. Ranking these importances at the respondent level, and then aggregating these ranks, shows that perceived health risks is most important for 18.4 percent of these respondents, second most important for 57.7 percent of these respondents, and third out of four for 21.9 percent of these respondents. Cumulatively, this shows that health risks are third or better in importance for 98.1 percent of these respondents. Overall, on average, price was most important, health risks second, taste third, and pack type fourth.[34] It is also useful to look at the ranks in the absence of price, since the other attributes represent what the consumer obtains by buying the cigarette in exchange for payment of the price. The relative rankings for the remaining three attributes are unchanged, i.e., health risks are, on average, most important, followed by taste and then pack type.[35] From these analyses, I conclude that perceived health risks are a significant contributing factor in the purchase decisions of 98.1 percent of "light" cigarette consumers who place a positive value on health risks.

[34] The average rank for price is 1.52, for perceived health risks 2.07, for taste 2.62, and for pack type 3.79. Standard errors for each average rank were in the range of 0.02 to 0.04.

[35] The average rank for perceived health risks is 1.37, for taste 1.80, and for pack type 2.83 in the absence of price. Standard errors for each average ranks are in the range of 0.030 to 0.042.

X. Calculations of Value of Perceived Reduced Health Risks Associated with "Light" Cigarettes

50. The HB partworth estimates can be used to determine the monetary value that consumers place on a having the lower perceived health risks associated with "light" cigarettes. I illustrate how to calculate the monetary value with two methods which indicate (1) how individual consumers value health risks and (2) how the market values health risks. The Willingness-to-Pay (WTP) method uses the relative partworths of price changes compared to the relative partworths of feature changes to impute the amount that each respondent would be willing to pay for lower health risks. I summarize these willingness-to-pay dollar values across respondents. The Market-Based method uses the partworths to predict how respondents would react in a "but-for" world in which I change the price level and the health risks level of a "light" cigarette. For each respondent the partworths (and the characteristics of the "but-for" world) imply choice probabilities. These choice probabilities, in turn, imply a market share for each "but-for" cigarette option. I use the Market-Based method to determine the price difference at which the market, as represented by the respondents, would be indifferent between a "light" cigarette with perceived health risks the same as "light" cigarettes and a "light" cigarette with perceived health risks the same as regular cigarettes. I also use the Market-Based method to gain insight on the price discount that would be needed to sell a cigarette with perceived health risks greater than regular cigarettes.

51. With the Willingness-to-Pay method, I first calculate the number of units of consumer utility that correspond to a one-percent change in price. For example, Exhibit G indicates that, on average, a change in price from the current pack price to 50 percent off is worth 47.9 units of consumer utility. This is equivalent to saying that, on average, 47.9 units of consumer utility

correspond to a price drop of 50 percent or that, based on the ratio of these two numbers, each unit of consumer utility corresponds to a price change of 1.04 percent.[36] For these same respondents, the number of units of consumer utility that are lost, on average, for going from perceived health risks the same as "light" cigarettes to perceived health risks the same as regular cigarettes, are 33.5 units of consumer utility. These calculations are illustrative. I perform these calculations for each respondent in the sample and then summarize the results of the respondent-level calculations. To complete the illustration, I consider a hypothetical situation in which all respondents have partworths equal to the average partworths. In this hypothetical case, the willingness-to-pay for a reduction in health risks from that perceived for regular cigarettes to that perceived for "light" cigarettes would be worth 34.8 percent (1.04 percent for each unit of consumer utility times 33.5 which represents the number of units of consumer utility gained by reducing the perceived health risk). This calculation for the average partworths is hypothetical and provided in this paragraph to illustrate the calculations that are made for each respondent. When partworths vary by respondent (heterogeneous respondents), the median willingness-to-pay can either increase or decrease relative to this hypothetical situation. When I complete this calculation for each respondent and take the median across respondents, I calculate a median willingness-to-pay of 47.3 percent.[37, 38]

[36] I choose the range of price discounts to consider to be consistent with the calculated WTP. In this illustrative case, the example price discount of 38.4% is in the range of 20% off to 50% off, thus it is appropriate to compare the partworths for price discounts of 38.5% and 50% off as opposed to using 20% off for the lower end of the range.

[37] I use the median as a statistic of central tendency rather than the mean because (1) willingness-to-pay is the ratio of two estimates, each with its corresponding variance, (2) the Bayesian methods provide complete estimates of the distribution of the partworths for each individual respondent (for the market-based method), (3) the median allows me to avoid extrapolation beyond the ranges of the price levels in the survey, (4) the median is a more robust statistic under these conditions, and (5) the median provides an accurate summary of central tendency. In the analyses in this report, the median is appropriately conservative because the WTP that I obtain from the median is less than the WTP that I obtain from the less-robust means. Calculations using other robust estimators yield answers that are similar to those obtained for the median. In addition, I performed sensitivity analyses using slightly different formulae to calculate WTP. For example, when I use formulae that take the non-linearity of the partworths into account, I obtain a median WTP of 45%.

52. The Market-Based method uses the HB partworths to simulate a market in which all respondents react to the same price. Such Market-Based simulations are often used by firms to simulate what would happen if a new product were introduced to a market or if an existing product changed its features. It is well known in the market research industry that the forecasts based on such Market-Based simulations are sufficiently accurate that firms routinely make decisions based on these simulations. The Market-Based method uses the complete population distribution of partworths (as estimated by HB) for each respondent. These partworths represent how each respondent will react to changes in features or price. Using the probability model (multinomial logit model) underlying HB I then use these partworths to predict the probabilities that each respondent would choose each of the service bundles from a defined choice set.

53. To calculate the value of a reduction in perceived health risks, I consider an alternative scenario in which there are only two alternative "light" cigarette options: one option with perceived health risks the same as "light" cigarettes and second option with perceived health risks the same as regular cigarettes. For each of these cigarette options I hold all other features constant at the same levels (type of pack, taste like the respondent's brand of "light" cigarette, and, for the first option, price that the respondent pays now for the respondent's brand of "light" cigarette). I then simulate markets in which the second cigarette option is offered at varying percentage changes in prices. This second cigarette option has all of the

[38] Because the survey allows the respondent to react to price changes up to 50%, any willingness-to-pay calculation that results in an estimate greater than 50% is an extrapolation and should be treated appropriately. In other words, if the extrapolations yield an estimated WTP greater than 50%, then we can be confident that the WTP is greater than 50%, but we cannot be confident in its exact value. For example, extremely high WTP results are due to extrapolation effects. Because the calculated median of 47.4 percent is less than 50 percent, the median is not affected by the details of the extrapolation and is, thus, robust with respect to extrapolation. To calculate this median we need only know that some consumers have WTPs that are above 50 percent. We do not need to calculate the exact values. On the other hand, if we were to use the mean WTP, the calculation of the mean would involve respondents whose estimated WTPs are based on extrapolation. In summary, the median is a robust measure that is, appropriately, conservative.

"light" cigarette features except that the perceived health risks are the same as regular cigarettes. I lowered the price of the second cigarette option until the market was indifferent between the two cigarettes. That is, I find the percentage price reduction for the second cigarette option such that half of the market chooses the first ("light" health risks) "light" cigarette option and half of the market chooses the second (regular cigarette health risks but lower priced) cigarette option. The price discount that is necessary for market indifference is the monetary value that the market places on the difference in perceived health risks.

54. There are two ways in which I calculate market share using the HB partworths. The difference is technical and requires a detailed understanding of CBC and the HB analysis. I attempt to provide a lay explanation here. Recall that HB provides the best estimate of a respondent's partworths. If I take the best estimate for each respondent, I obtain a market simulation based on the "point" estimate for each respondent of that respondent's partworths. This "point" estimate gives a "point" estimate for each respondent's utility for each profile in the market. The respondent "chooses" the profile with the highest utility. Using this method I find the price at which the market is indifferent. I provide the results of this calculation in Exhibit H as labeled by "First Choice Simulation." The words, "first choice," refer to the fact that I use the best estimate and calculate the respondent's first-choice cigarette option, i.e., which cigarette option has the highest overall utility, based on the sum of the option's partworths for that respondent. The second method uses detailed information (from the HB estimation) on the unobserved features and randomness in consumer choice. In accord with the logit model, utility is based on the best estimate of the partworths plus a random error consistent with the estimation. I repeatedly sampled from the distribution of this error and compute the respondent's first choice cigarette option (based on appropriate probabilities) for

each sample that is drawn. This is known as the "Randomized First Choice Simulation." By using two procedures to obtain a market-indifference price, I am able to determine whether or not the two procedures provide similar estimates for the data obtained from the survey described in this report. The variation in the two procedures provides insight on the possible range of estimates.

55. The results of the Market-Based simulations are given in Exhibit H. The Market-Based method estimates that the value of the change in perceived health risks from the same as regular cigarettes to the same as "light" cigarettes is approximately 47.1 percent of the price per pack by the first-choice method and 39.8 percent of the price per pack by the randomized first-choice method. Thus, using the Market-Based method I estimate that the value of the change in perceived health risks from the same as regular cigarettes to the same as "light" cigarettes is in the range of 39.8 to 47.1 percent of the price per pack. Because these estimates are approximately equal to those obtained by the Willingness-to-Pay method (47.3 percent), it confirms my confidence that the estimated range is accurate.

56. The lowest level of price used in the Choice-Based Conjoint analysis was "50% LESS than what you pay now" and the highest level of price was "50% MORE than what you pay now." Thus, the data can be used confidently to calculate any willingness-to-pay that is between 50 percent less and 50 percent more than what the consumer now pays. If the willingness-to-pay calculation results in a number that is larger than 50 percent more (or less than 50 percent less), then we can be confident that the willingness to pay is larger than 50 percent more (or less than 50 percent less). However, because the calculation is based on extrapolation, we should be extremely cautious when interpreting the exact value. I, therefore, use methods of analysis that avoid extrapolation. In particular, I calculate that more than 75 percent of the

consumers would be willing to pay more than 50 percent of the price per pack of their cigarettes to decrease health risks from greater than regular cigarettes to health risks the same as "light" cigarettes.[39]

57. The data and analyses in this report can also be used confidently in a Market-Based Simulation to examine markets in which a cigarette is priced at 50 percent less that what the consumer now pays. I consider a scenario in which there are only two alternative "light" cigarette options: one option with perceived health risks the same as "light" cigarettes at a price that the respondent pays now for the respondent's brand of "light" cigarette and a second option with perceived health risks greater than regular cigarettes but with a price that is 50 percent less than the respondent pays now for the respondent's brand of "light" cigarette. For each of these cigarette options I hold the other features constant at the same levels (type of pack and taste like the respondent's brand of "light" cigarette). Using the First Choice Simulation, I estimate that the cigarette with health risks the same as "light" cigarettes would obtain a 75.4 percent market share even though the second higher-health-risks cigarette was offered at a price 50 percent less than the respondent pays now. Using the Randomized First Choice Simulation, I estimate the market share of the cigarette with health risks the same as "light" cigarettes to be 72.3 percent. Thus, in order for the market to be indifferent between the two types of cigarettes, the cigarette with health risks greater than regular cigarettes would have to offer substantially more than a 50 percent discount per pack.

[39] The estimates are 75.4% and 75.3% for simple and non-linear price response calculations, respectively.

XI. Overview of Methodology – Experimental Time Study (Draft)

58. In a Memorandum and Order – Schwab v. Philip Morris USA, Inc., 04-CV-1945, October 6, 2005, Judge Jack B. Weinstein, Senior United States District Judge, Eastern District of New York, wrote:

> "Based on surveys, other evidence and extrapolation, plaintiffs' experts might develop a model showing how many smokers of "light" cigarettes were ignorant of the alleged fraud in each relevant year."

59. Plaintiffs requested that I attempt to design and implement a survey research study to provide evidence with which to assess "how many smokers of "light" cigarettes were ignorant of the alleged fraud in each relevant year." If, based on initial tests, smokers could provide reliable and valid responses to retrospective questions, I would then undertake a more comprehensive study as described below.

60. In order to provide survey evidence that might inform such a model, I designed an experimental study that I call the "Experimental Time Study." I designed the Experimental Time Study to determine whether it was feasible to obtain accurate survey evidence about two issues. The first issue is the distribution of the years in which current and former "light" cigarette smokers first changed their beliefs about the relative heath risks of "light" cigarettes versus regular cigarettes, if at all. The second issue is the distribution of the years in which current and former "light" cigarette smokers first knew or suspected that cigarette companies had not told the truth about the relative heath risks of "light" cigarettes versus regular cigarettes, if at all.

61. It is, based on my experience, more difficult for survey respondents to recall the year in which an event occurred than it is to answer questions such as those asked in the Conjoint

Study. Before valid results can be obtained, I first need to know whether the events of interest are the types of events that respondents can remember accurately and with confidence. Potential errors that can occur include forgetting that an event occurred (omission), or incorrectly moving an event forward or backward in time (telescoping).[40] While there are many examples of omission and telescoping in the marketing research literature, these potential errors do not apply equally to all events. There are conflicting results in the research about the effects of different methods used to deal with these errors. For example, on pages 88-89 of their text on *The Psychology of Survey Response*, Roger Tourangeau, Lance J. Rips, and Kenneth Rasinski state:[41]

> "Respondents' uncertainty about the start of the reference period implies that they may mistakenly report events that happened before it or mistakenly omit events that happened during it in answering a survey question. The first type of mistake goes by the name of *forward telescoping*, and the second is called *backward telescoping*…we note simply that methods that appear to reduce telescoping call the respondents' attention to personal facts that can serve as a boundary for the reference period." (Tourangeau et. al., pp. 88-89)"

62. Recognizing that it was necessary that respondents be able to remember the events of interest accurately and with confidence, I sought first to test the potential methodology with respondents who would be most likely to provide accurate judgments as to the time at which they changed their beliefs about the two issues of interest. I chose to focus on those respondents

[40] Tanur, Judith M. (1992). *Questions About Questions: Inquiries into the Cognitive Bases of Surveys.* Russell Sage Foundation, New York, page 88.
[41] Roger Tourangeau, Lance J. Rips, and Kenneth Rasinski (2000), *The Psychology of Survey Response*, (New York, NY: Cambridge University Press), pages 88-89.

who, at some time now or in the past, reported a belief that smoking "light" cigarettes provided substantially less health risk than regular cigarettes. I chose these respondents because, a priori, I believed that respondents who once perceived a substantial difference and now perceive less difference would be more likely to be able to remember when they changed their beliefs. The switch from a belief that there was once a substantial difference in health risk would be more likely to be salient and more likely to be easily placed in time. If I could obtain valid and reliable answers from respondents who once held a belief about a substantial difference, then I could be more confident in a larger study of all respondents who reported a belief that "light" cigarettes provided less health risk than regular cigarettes. If respondents were not confident in their answers or if there was evidence of telescoping or related phenomena, then I would seriously evaluate whether to undertake a study of all respondents who perceived, now or in the past, that "light" cigarettes provided less health risk than regular cigarettes.

63. The potential errors of omission and telescoping do not apply equally to all events. Some events are more difficult to place in time with reasonable accuracy; other events are less difficult to place in time. For example, if a respondent is asked to remember the birth of the respondent's first child (if applicable), then most respondents could give the year that the birth occurred, or failing that, a range of years in which the birth occurred. On the other hand, if the respondent is asked to remember his or her last purchase of furniture polish (if such a purchase was made), the exact placement of that event in time may be more difficult. The birth of one's first child is much more memorable than the last purchase of a consumer packaged good. Even recent events may be difficult to recall if they are part of a daily (or weekly) routine that varies slightly. For example, a respondent might be better able to recall

what he or she ate for dinner the previous night, but may have more difficulty remembering what he or she ate for dinner one week ago. On the other hand, if the routine follows a rigorous schedule, the event will be easier to recall. If the respondent eats the same thing for breakfast every morning, it is easier for the respondent to recall what he or she ate for breakfast one week ago. Research suggests that the difficulty of recalling events is related to how long ago they occurred and the saliency of the event or behavior bring measured.[42]

64. People tend to have a much better ability to recall the date of "landmark" events that are of great personal importance. Many people, for example, can recall accurately the year they were born, or the year they graduated from high school or college, or the year they got married (if they are married or were once married).[43] These events can be used to help place other events in time if the other events occurred at roughly the same time. For example, many people recall where they were on the morning of September 11, 2001 even though they might have more difficulty remembering where they were on September 11, 2005.

65. The two events that I sought to determine whether a respondent could recall were (1) when a respondent came to change his or her beliefs about the health risks of "light" cigarettes and (2) when a respondent came to change his or her beliefs about whether cigarette companies were or were not telling the truth about health risks regarding "light" cigarettes. A priori, I believed that the ability to recall these two events would be less than the ability of a respondent to recall where he or she was on the morning of September 11, 2001 and more than the ability of a respondent to recall the timing of his or her last purchase of furniture polish (if such a purchase occurred and if that purchase was not recent). I could best assess respondents' abilities to recall the two events of interest with an actual experimental survey.

---

[42] *Ibid.*

[43] Tourangeau, Roger, Lance J. Rips and Kenneth Rasinski (2000). *The Psychology of Survey Response*. Cambridge University Press, Cambridge, UK, p. 114.

66. To help respondents recall the two events of interest and to give the Experimental Time Study the best chance of succeeding, I used respondents' memories for landmark events of a personal nature to help them place in time when their beliefs changed relative to the health risks "light" cigarettes and relative to whether or not they were being misled by the cigarette companies. To further reduce the likelihood of omissions and telescoping, respondents were encouraged to take their time, and offered aids for recall, including a calculator (to assist in converting ages to approximate calendar years) and timelines of well-known events. In addition, I followed all appropriate scientific survey research methods as described in subsequent paragraphs.

XII. Questionnaire Development – Experimental Time Study (Draft)

67. Qualitative Interviews. At my direction, AMS conducted a total of 9 interviews from October 27th through November 4th, 2005.

68. The pretests suggested that these landmark events helped respondents to remember the years that they changed their perceptions of the health risks of "light" cigarettes, if they changed their perceptions, and the years that they changed their perceptions of whether or not the cigarette companies were misleading them (the respondents), if they changed their perceptions. The pretests suggested that some respondents might wish to use a calculator to identify the year that corresponded to their age at the time an event occurred. For example, a respondent might remember that they were 34 years old when a parent died, but needed a calculator to calculate the calendar year. Based on the pretests, I developed a custom calculator that enabled respondents to enter their current age, the age they believed they were when the event in question occurred, and obtain the (approximate) year when they were that age. Respon-

dents were also provided with buttons that they could, at their discretion, click on to view timelines of world events and films. These timelines contained the dates of memorable events, such as the Space Shuttle Challenger disaster (1986) and the premier of the original *Star Wars* movie (1977). Content for the timelines was obtained from the Encyclopedia Britannica website (www.britannica.com) and the Wikipedia web site (en.wikipedia.org). Wikipedia is a free Internet-based encyclopedia maintained by a large community of users. At the end of the pretests, most respondents expressed the belief that the cues were helpful and relevant.

69. The questionnaire was programmed into a web-based software system designed for administering and analyzing such questionnaires.[44] Examples of the final questionnaire that respondents were asked to complete are shown in Exhibit J. Recall that respondents answered these questions via their computers. Exhibit J contains reproductions of example computer screens shots. Following accepted scientific methodology, some questions, such as the number of packs of cigarettes smoked per day, included appropriate randomization of the response scale to avoid order effects. I have also provided a Microsoft Word version of the questionnaire in Exhibit J that includes randomization and skip-pattern instructions. The computer program that was used to generate the screens has been provided as part of the discovery materials.

XIII. Pretesting the Questionnaires – Experimental Time Study (Draft)

70. Following the qualitative pretests, the web-based questionnaire was pretested with 20 additional respondents on November 21st through November 28th, 2005. Following standard procedures, no pretest responses were included in the final sample. Based on the pretest, I

[44] The survey was programmed in ASP, with a SQL Server database, both of which are well-known and widely used software systems for these types of applications.

was satisfied that the respondents understood the questions.

XIV. Identifying the Sample – Experimental Time Study (Draft)

71. For the Experimental Time Study, potential respondents were selected at random from Greenfield Online's database and sent an invitation (Exhibit K) to go to a special website to complete the survey. Experimental Time Study respondents who qualified for study and completed the survey were given an entry into Greenfield Online's Home for the Holidays Sweepstakes, with a chance to win $4,000 in cash prizes. In my experience, such incentives increase response rates but do not bias any of the responses to the questions in the survey.

72. In order to qualify for the survey, respondents were screened to ensure that they were current or former "light" cigarette smokers. A total of 1,026 respondents completed the survey beginning on November 28th, 2005 and ending on December 6th, 2005. The completion rate after the respondent was qualified as a current or former "light" cigarette smoker was 95.4 percent. Details are provided in Exhibit L.[45]

73. To ensure a nationally representative sample of respondents, quotas were set so that the sample would match the national data on Census region, sex, age, and household income. These respondents were allocated to the seventy-two quota groups (four Census Regions crossed with two levels of sex, three levels of age, and three levels of household income, 72 =

[45] For the Experimental Time Study, the incidence rate was 26.8 percent, the response rate was 5.7 percent, the completion rate (after qualified as a current or former light cigarette smoker) was 95.4%, and the net response rate was 5.4%. The incidence rate for the Experimental Time Study is larger than that for the Conjoint Study because the Experimental Time Study include current and former "light" cigarette smokers while the Conjoint Study included only current "light" cigarette smokers. The response rate (and net response rate) for the Experimental Time Study is likely lower than that for the Conjoint Study because respondents were not given the same incentives to participate. The comparative response rates suggest that an entry is Greenfield Online's Home for the Holidays Sweepstakes is perceived by respondents to be less of an incentive than a $5 participation incentive. Because the focus of the Experimental Time Study was on determining whether or not respondents could provide retrospective estimates of the time at which the two events of interested occurred, this lower response rate is acceptable. Where I to expand the Experimental Time Study, I would provide the greater incentive to achieve a response rate in line with the response rate obtained in the Conjoint Study.

4x2x3x3).[46] In order to identify 4,005 potential respondents, 4,489 potential respondents were screened. Exhibit L shows the number of respondents in each group and the number completing the questionnaire. The completed interviews match closely the screening quota and the census categories.[47]

74. The target universe from which I sampled respondents for the Experimental Time Study was wider than the target universe from which I sampled respondents for the Conjoint Study because the target universe for the Time Survey includes former smokers of "light" cigarettes in addition to current smokers of "light" cigarettes. After answering the demographic quota allocation questions, respondents were asked whether they were current or former cigarette smokers.[48] The allocation and screening questions yielded a representative sample of 1,026 current or former "light" cigarette smokers.[49] These included 584 current and 442 former smokers of "light" cigarettes.

XV. Focus on Substantial Changes – Experimental Time Study (Draft)

75. Identifying respondents with substantial changes in beliefs. In the Conjoint Study I measured the current stated beliefs of respondents about the health risks of "light" cigarettes. To measure those beliefs I used a numerical scale on which respondents could indicate their perceptions of the health risks of "light" cigarettes relative to their perceptions of the health risks of

[46] The respondents must be 18 years of age to participate.
[47] There is no reason to expect that the penetration of current and former "light" cigarette consumers will match the distribution of census regions and/or the distribution of age, sex and household income. The screening ensures we start out with a representative sample. The final sample reflects the distribution of current and former "light" cigarette consumers as observed.
[48] The distributions of current smokers of "light" cigarette smokers did not vary significantly between the Conjoint Study and the Experimental Time Study on sex ($\chi^2 = 0.020$, p = 0.88) or income ($\chi^2 = 0.117$, $p = 0.94$) The distributions did vary on age ($\chi^2 = 19.25$, p = 0.00). The younger ages in the Experimental Time Study might be due to the difference in incentives between the Experimental Time Study and the Conjoint Study. Were I to expand the Experimental Time Study, I would provide the greater incentive in order to achieve a response rate and demographic distribution in line with that obtained in the Conjoint Study.
[49] Because the incidence of smoking may vary by the demographic categories, we do not expect the demographic distribution of current and former "light" cigarette smokers to match the distribution of all potential respondents (current and former smokers + non-smokers).

regular cigarettes. The numerical scale enables respondents to provide a detailed, fine-grained opinion about these relative health risks. As indicated in Exhibit D, respondents are asked to provide a numerical value between 0 and 150 to indicate their beliefs. A value of zero represents a harmless cigarette, a value of 100 represents the perceived health risk of smoking a regular cigarette, and a number larger than 100 represents higher perceived health risks than smoking a regular cigarette. This scale was appropriate for the Conjoint Study because my focus in that study was on the issue of whether or not respondents value changes in health risk and, if they value such changes, the monetary value of such changes.

76. In the Conjoint Study, 37.2% of the respondents held beliefs at the time of the survey that the health risks of "light" cigarettes were exactly equal to those of regular cigarettes (an index of 100).[50] A small number, 1.3%, answered with an index that was greater than 100, indicating that they believed that "light" cigarettes had greater health risks than regular cigarettes. The remaining 61.5% of the Conjoint Study respondents gave an answer that indicated that they believed that "light" cigarettes had less health risks than regular cigarettes. Of these 61.5% of the respondents, 47.3% of the Conjoint Study respondents answered with an index between 75 and 100 indicating that they believed that "light" cigarettes had somewhat less health risk than regular cigarettes. A smaller percentage of Conjoint Study respondents,

[50] In the Conjoint Study I did not ask respondents to indicate if they had held different beliefs in the past. Furthermore, because I used the aggregated scale in the Experimental Time Study rather than the numerical scale in the Conjoint Study, I cannot estimate the percentage of respondents who would have indicated that they changed their beliefs about the health risk of "light" cigarettes. For example, we know that 61.5% of the Conjoint Study respondents currently believe that "light" cigarettes have less health risk than regular cigarettes. From the Experimental Time Study we estimate that, of those respondents who answer "same" or "more," 29.8% changed their beliefs relative to the aggregated scale. Thus, I can estimate that, at minimum, 61.5% + (29.8%)(37.2%) = 73.0% held a belief at some time that "light" cigarettes had less health risk than regular cigarettes. However, this percentage underestimates the percent of customers who held that belief because it does not include those who may now answer with an index of 100 and would have, in the past, answered with an index between 75 and 100. Given that over three times as many respondents currently answer with an index between 75 and 100 compared to those who answer with an index less than 75, the 73% could be substantially higher.

14.2%, answered with an index of less than 75, indicating that they believed that "light" cigarettes had substantially less health risks than regular cigarettes.

77. A priori, I believed that respondents would be better able to remember the time that they changed their beliefs if they held strong beliefs in the past and now held a belief that was different. Thus, in the Experimental Time Study I used a more aggregated health-risk scale that would be simple to administer and would focus on such respondents.[51]

78. Qualitative interviews and pretests of the Experimental Time Study indicated that many respondents could provide confident answers with respect to changes in their beliefs relative to these more aggregate categories of health risk. Specifically, in the Experimental Time Study I used a three-point rather than a 150-point scale. Respondents in the Experimental Time Study were asked if they personally believe that smoking light cigarettes has MORE health risks than smoking regular cigarettes, THE SAME health risks as smoking regular cigarettes, or LESS health risks than smoking regular cigarettes. More aggregated categories mean that we gather less detailed information. An analogy might be if we attempted to measure how far west or east cities are in the United States. The exact "west-ness" or "east-ness" can be measured with longitude. However, we might decide to ask a respondent whether the city is in the western US, the eastern US, or in-between. A respondent might place both St. Louis and Chicago "in-between," even though Chicago is east of St. Louis and both are east of the mid-point of the contiguous US.

[51] Because the focus of the Experimental Time Study was to determine whether or not respondents could provide accurate and confident estimates of the years in which the events of interest occurred, the aggregated scale is not intended to be a discriminating measure of respondents' perceptions of the relative health risks of "light" cigarettes vs. regular cigarettes. To measure these perceptions, it is better to rely upon the more accurate assessments in the Conjoint Study. An alternative would have been to use the original 0-150 scale and devise a corresponding screening algorithm.

79. To determine the comparability of the Conjoint Study scale and the Experimental Time Study scale, I compared respondents' perceptions of health risks using the detailed scale in the Conjoint Study and the aggregated scale in the Experimental Time Study. In the Experimental Time Study, of those offering an opinion,[52] 16.0% of the respondents reported that they believed that smoking "light" cigarettes had less health risks than smoking regular cigarettes. This is comparable to the 14.2% of the respondents in the Conjoint Study who provided a response less than 75 of 100. In the Experimental Time Study, of those offering an opinion, 81.9% of the respondents reported that they believed that smoking "light" cigarettes had the same health risks as smoking regular cigarettes. This is comparable to the 84.5% of the respondents in the Conjoint Study who provided a response in the range of 75 to 100.[53] In the Experimental Time Study, of those offering an opinion, 2.0% of the respondents reported that they believed that smoking "light" cigarettes had more health risks than smoking regular cigarettes. This is comparable to the 1.3% of the respondents in the Conjoint Study who provided a response in the range of 100 or more. Using a χ^2 statistic to compare the responses in the Conjoint Study to those in the Experimental Time Study, I find no significant differences ($\chi^2 = 0.73$, $p = 0.70$) between these two sets of three aggregated categories. I also compared current and former smokers' responses in the Experimental Time Study. There were no significant differences ($\chi^2 = 2.42$, $p = 0.49$).[54] Based on these statistical analyses I conclude that the two-scales were comparable in the sense that, on average, those respondents who answered "less" were comparable to those respondents who would have answered with an index of less than 75 in the Conjoint Study. One consequence of this scale is that

[52] 36 respondents, or 6.2% of the 1,026 in the total sample, answered "I don't know / I am not sure."
[53] 47.3% gave a response from 75 to 99 and 37.2% gave a response of 100.
[54] The percentages (including all responses) for current smokers were 15.1%, 76.9%, 1.9%, and 6.2% for less, the same, more, and don't know/not sure, respectively. The percentages for former smokers were 14.3%, 76.5%, 3.4%, and 5.9%, respectively.

those respondents who would have answered with an index between 75 and 100 are now merged with those respondents who would have answered with an index of 100.[55]

XVI. Confidence Measurement – Experimental Time Study (Draft)

80. So that I might ascertain whether or not respondents were confident in the answers that they provided about the time at which they changed their beliefs, I measured their confidence in their answers. Respondents were asked questions about their confidence in the year(s) they had estimated for each of the events. Respondents gave answers either for zero, one, or two estimates as described by the previous paragraphs. Respondents were only asked to indicate their confidence for the events for which they estimated a year or years. Respondents were reminded of the event and their estimate and were asked, "Using the response scale below, please indicate how confident you feel about THE YEAR(S) you told us. Selecting a lower number indicates you are LESS confident and selecting a higher number indicates you are MORE confident." The responses to these confidence questions were given on a seven-point scale using radio buttons. "Not Confident" was given a value of one, and "Confident" was given a value of seven. See Exhibit J for screen shots of these questions. The answers to these questions about confidence also enable me to compare the answers by confident respondents to the answers by respondents who are less confident.

[55] This scale-compression phenomenon is documented in the literature on the psychology of survey response. By using aggregated scales, more respondents might choose the middle category, thus, any respondent who choose the lower category, "less," for one of his or her responses might be more likely to be able to recall when he or she changed his or her belief. See, for example, Tourangeau, et. al., ibid., p. 232.

XVII. Survey Administration – Experimental Time Study (Draft)

81. After initial screening for representativeness and for whether or not they were current or former "light" cigarette smokers, respondents were asked additional questions. Respondents were asked how long they had been smoking, what type of cigarette they had smoked primarily when they began smoking, and how may years they had started to smoke primarily "light" cigarettes. They were asked how many packs per day they smoke(d) and were asked to indicate which brand of "light" cigarette is/was their primary brand. Respondents were asked for which brands of "light" cigarettes they had seen or heard advertising in the past month. The advertising question was included to help disguise the purpose of the survey and, hence, minimize any demand artifacts.[56]

82. Respondents then indicated their personal beliefs about the health risks of "light" cigarettes using the aggregated three-point scale. Respondents who indicated that their personal belief was that smoking "light" cigarettes had "more health risk than regular cigarettes" or "the same health risk as regular cigarettes," were asked a question to ascertain when they came to believe these statements. Respondents were asked to select one of the following statements regarding this belief: "I have ALWAYS believed this," "I have NOT ALWAYS believed this," or "I don't know/I am not sure." If respondents answered that they "have always believed this" or that they "don't know or are not sure," they skipped to the next set of questions about whether or not they were misled by the cigarette companies.

83. If respondents had believed that smoking "light" cigarettes had more health risks than smoking regular cigarettes (on the aggregated scale) and that they had NOT ALWAYS believed this, they were asked, "Prior to your changing your belief about the health risks of smoking

[56] There was no need to ask respondents about their preferences for the type of pack and taste in the Experimental Time Study. Thus, the Microsoft Word version of the questionnaire skips from Qs5 to Q8. Recall that the respondents do not see any numbering scheme in the web-based survey.

light cigarettes having MORE health risks than smoking regular cigarettes, did you believe that smoking light cigarettes had "the SAME health risks as smoking regular cigarettes," "LESS health risks than smoking regular cigarettes" or "I don't know/I am not sure." If respondents had believed that smoking "light" cigarettes had the same health risks as smoking regular cigarettes and that they had NOT ALWAYS believed this, they were asked, "Prior to your changing your belief about the health risks of smoking light cigarettes having the SAME health risks as smoking regular cigarettes, did you believe that smoking light cigarettes had "MORE health risks as smoking regular cigarettes," "LESS health risks than smoking regular cigarettes" or "I don't know/I am not sure."

84. If respondents had changed their beliefs and if their prior belief was that "light" cigarettes had less health risks than regular cigarettes (on the aggregated scale), respondents were asked, "do your best to remember IN WHAT YEAR your belief about the health risks of light cigarettes CHANGED." These respondents were given a list of landmark events and buttons they could click to access world events and movies timelines to help them remember (see Exhibit J). A calculator was also provided to help convert ages when events occurred into approximate years. Respondents were given as much time as they needed to remember when they had changed their beliefs. They could access the timelines and calculator as often as they wanted. When they were ready they clicked the "NEXT" button to continue.

85. On the next screen, respondents were given the option of entering a particular year, entering a time interval (a range of years), or selecting that they had "no idea and nothing will change that." Depending upon their choice about how they wanted to answer this question, they were given the opportunity to either enter a year or a range of years. The estimates were constrained by the survey software to be contained in the interval from 1971 through 2005.

The lower bound of 1971 was chosen because that is the year, I am informed, in which "light" cigarettes were first introduced.

86. In the next section of the survey, respondents were asked questions about their beliefs about whether or not they were misled by the cigarette companies. Specifically, the objective of this part of the survey was to assess respondents' beliefs as to when (if ever) they first came to know or suspect that cigarette companies had not told the truth about the relative health risks of "light" cigarettes vs. regular cigarettes.

87. Following accepted procedures, respondents were first asked if they had ever seen or heard any information about cigarette companies telling the truth or not telling the truth about their products. Respondents were given the option of responding that they had seen or heard information about cigarette companies "TELLING THE TRUTH about their products," "NOT TELLING THE TRUTH about their products," BOTH TELLING THE TRUTH AND NOT THE TRUTH about their products," that "I have not heard anything," or "I don't know/I am not sure."

88. If respondents had seen or heard information about cigarette companies either telling the truth about their products or respondents had not heard anything, they were skipped to the questions about their confidence in their answers (following the indicated skip pattern).

89. If respondents had ever seen or heard information about cigarette companies not telling the truth, or both telling the truth and not the truth, or they did not know/were not sure, they were asked which "subject areas have you ever seen or heard any information about cigarette companies telling the truth or not telling the truth about their products." Respondents could indicate that they had seen or heard information about cigarette companies telling the truth, not telling the truth, both the truth and not the truth, or nothing with respect to four issues:

how addictive cigarettes might be, health risks of "light" cigarettes vs. regular cigarettes, cigarette ingredients, and cigarettes and lung cancer. These issues were rotated randomly in the survey to minimize order bias.

90. If respondents answered that they had seen or heard information about cigarette companies not telling the truth, or telling both the truth and not the truth with respect to either the health risks of "light" cigarettes vs. regular cigarettes or cigarettes and lung cancer, they continued to the question that asked them to estimate the year(s) in which they had "first suspected or became aware that cigarette companies had not told the truth about the health risks of light cigarettes." Otherwise, they skipped to the questions about their confidence in their answers (following the indicated skip pattern).

91. Respondents who had seen or heard information about cigarette companies not telling the truth or telling both the truth and not the truth with respect the health issues were asked "to do your best to remember "IN WHAT YEAR YOU FIRST SUSPECTED OR BECAME AWARE that cigarette companies HAD NOT TOLD THE TRUTH ABOUT THE HEALTH RISKS of light cigarettes." Following parallel methods to those used to assess respondents' changes in beliefs about the health risks of "light" cigarettes, qualified respondents were given the same set of landmark events, buttons that would lead to timelines of world events and films, and a calculator with which to translate their age at the time of an event into an approximate calendar year. Respondents were given the options of providing their answers in terms of a single year, a time interval (a range of years), or selecting that they had no idea and nothing would change that. Respondents who indicated they had no idea and nothing would change that were not asked to estimate the year in which they first suspected or be-

came aware that cigarette companies had not told the truth about the health risks of light cigarettes.

92. Finally, respondents were asked questions about their confidence in the year(s) they had estimated for each of the events.

XVIII. Analysis – Experimental Time Study (Draft)

93. Of the 1,026 respondents in the Experimental Time Study, 207 respondents (20.2%) attempted to give a year or a range of years when they first changed their beliefs about the relative health risks of smoking "light" vs. regular cigarettes. Of these, 126 respondents (61%) gave their estimate in the form of a single year, and 81 (39%) gave their estimate in the form of a range of years.

    a. Of the 126 respondents who gave a single-year estimate, the earliest was 1974 (1 respondent), the latest was 2005 (6 respondents), and over 50% of the responses were 1998 or later.

    b. Of the 81 respondents who opted to give their estimate in the form of a range, Exhibit M gives the low and high ends of the ranges. For the lower end of the ranges, the earliest year was 1975, the latest was 2004, and over 50% of the responses were 1995 or later. For the higher end of the ranges, the earliest year was 1976, the latest was 2005, and over 50% of the responses were 2000 or later.

94. The full distributions for the single year and ranges of years estimates, as well as the distribution of the lengths of the ranges, are given in Exhibit M. The distributions in Exhibit M reflect the answers provided by the respondents. As an example, in Exhibit N, I provide one method by which respondents can be apportioned to specific years by assigning each respon-

dent to the midpoint of the range that the respondent provided.[57]

95. Many respondents did not express full confidence in their answers. Only about a third of the 207 respondents who gave a year estimate in Q13 or Q14 (33.8%) were fully confident in their estimate of the year in which they changed their beliefs. However, 84.5% of these 207 respondents answered either a 5, 6, or 7 (out of 7) to question Q20, "please indicate how confident you feel about THE YEAR(S) you told us." The full distribution is given in Exhibit M. Exhibit N compares the distribution of respondent answers from those respondents providing a 5, 6, or 7 as compared to those respondents providing a 1, 2, 3, or 4.

96. I am informed that the public health community did not reach a consensus that "light" cigarettes had the same health risk as regular cigarettes until approximately 2001. If this is the case and if respondents did not have a means to form these beliefs prior to 2001, then it appears that respondents are either telescoping their responses to earlier dates, remembering other events, or some other reporting error. Furthermore, these effects appear to be greater for less confident respondents.

97. Of the 1,026 respondents in the Experimental Time Study, 371 respondents (36.2%) answered that they had "seen or heard information about the cigarette companies telling "NOT THE TRUTH" or "BOTH THE TRUTH AND NOT THE TRUTH" about the health risks of "light" cigarettes vs. regular cigarettes, and they answered that they could tell us the year or give us a time interval in Q17, in which they were asked "CAN YOU TELL US THE YEAR IN WHICH YOU FIRST SUSPECTED OR BECAME AWARE that cigarette companies HAD NOT TOLD THE TRUTH ABOUT THE HEALTH RISKS of light cigarettes?"

98. Of these 371 respondents who could give a year or a range of years in which they first suspected or became aware that cigarette companies had not told the truth about the health risks

[57] The midpoint is also the mean (and the median) of any symmetric probability distribution defined on the range.

of "light" cigarettes, 197 respondents (53% of the 371 respondents giving estimates) answered that they could give their estimate in the form of a single year, and 174 respondents (47% of the 371 respondents giving estimates) said they could give their estimate in the form of a range of years.

a. Of the 197 respondents giving a single-year estimate, the earliest year was 1971 (2 respondents) and the latest year was 2005 (2 respondents). Over 50% of the responses were 1995 or later.

b. For 174 respondents who chose to give their estimate in the form of a range of years, Exhibit O gives the distribution of the ranges of the these estimates. For the lower end of their estimates, the earliest year was 1971, the latest year was 2004, and over 50% of the responses were 1990 or later. For the higher end of their estimates, the earliest year was 1973, the latest year was 2005, and over 50% of the responses were 1996 or later.

99. The full distributions for the single year and ranges of years estimates, as well as the distribution of the lengths of the ranges, are given in Exhibit O. In Exhibit P, I provide one method by which respondents can be apportioned to specific years by assigning each respondent to the midpoint of the range that the respondent provided.

100. Fewer respondents (24.6%) expressed full confidence in the year in which they first suspected or became aware that cigarette companies had not told the truth about the health risks of light cigarettes than those that expressed full confidence in their estimate of the year in which they changed their beliefs about health risks. However, of the 371 respondents making such an estimate, 270 (72.8%) of the respondents answered either a 5, 6, or 7 (out of 7) to question Q21, "please indicate how confident you feel about THE YEAR(S) you told us."

The full distribution is given in Exhibit O.

101. I am informed that information about whether or not the cigarette companies had been misleading the public about the health risks of "light" cigarettes was not available to the public until approximately 1998 at the earliest. If this is the case and if respondents did not have a means to form these beliefs prior to 1998, then it appears that respondents are either telescoping their responses to earlier dates, remembering other events, or some other reporting error. Furthermore, these effects appear to be greater for less confident respondents.

XIX. Conclusions – Experimental Time Study (Draft)

102. In some instances respondents can give accurate, reliable, and valid estimates as to the time at which an event took place. In other instances, such judgments are difficult. Estimates of the years (1) in which a respondent changed his or her belief about health risk and (2) in which a respondent first came to believe that he or she was being misled by cigarette companies, may be either easy or difficult (or somewhere in between) for a respondent to provide. Theory alone cannot answer this question. Therefore, I designed a study to determine whether these estimates were feasible for respondents to provide. The study followed careful scientific procedures in the selection of the sample, the design of the questionnaire, the administration of the study, and the analysis of the data. Furthermore, for this experimental study, I focused on those respondents who once perceived a substantial difference between the health risks of "light" cigarettes and the health risks of regular cigarettes.

103. Based on the Experimental Time Study, I do not have confidence that most respondents can provide accurate estimates of the years in which the two events of interest occurred.

a. Not all respondents were willing to answer the time questions.

b. Of those respondents who answered the time questions, most did not express full confidence in their answers. Many were not confident in their answers (a confidence scale value of 4 or less).

c. There was evidence of telescoping their responses to earlier dates, remembering other events, or some other reporting error.

104. Because I do not have confidence in the time estimates obtained by the Experimental Time Study, I do not intend to carry out the Experimental Time Study for all respondents, including those who can express their perceptions of health risk on the numerical scale with an index between 75 and 100.

XX. Summary of Conclusions - Overall

105. The scientific methodology used to design, execute, and analyze the Conjoint Study in this report is sound, reliable, and valid. The results can be relied upon to draw inferences about whether perceived health risks are a significant contributing factor in consumer decisions to smoke "light" cigarettes and what proportion of "light" cigarette-smoking consumers relied on health risks as a significant contributing factor. The results can also be relied upon to draw inferences about the value to consumers of the change in perceived health risks from the same as regular cigarettes to the same as "light" cigarettes. I illustrate how the results can be used with methodologies based on (1) consumers' willingness-to-pay for reduced health risks and (2) the market's valuation.

106. Based upon an examination of the partworths for perceived health risks, I find, to a reasonable degree of scientific certainty, that 69.7 percent of "light" cigarette consumers place a statistically significant positive value on perceived health risks, and only 1.1 percent place a statistically significant negative value on perceived health risks. In accordance with these as-

sessments and measurements, 90.1 percent of "light" cigarette consumers place a positive value on this change in perceived health risks.

107. Based upon an examination of feature importance, I find, to a reasonable degree of scientific certainty, that perceived health risks are a significant contributing factor in the cigarette purchase decisions of 98.1 percent of "light" cigarette consumers who place a positive value on perceived health risks.

108. Based on the Willingness-to-Pay method, I estimate that the median value of the change in perceived health risks from the same as regular cigarettes to the same as "light" cigarette is 47.3 percent of the price per pack.

109. Based on the Market-Based method, I estimate that the market value of the change in perceived health risks from the same as regular cigarettes to the same as "light" cigarette is between 39.8 percent and 47.1 percent of the price per pack.

110. Based on the Willingness-to-Pay method, I estimate that more than 75 percent of the consumers would be willing to pay more than 50 percent of the price per pack to decrease health risks from greater than regular cigarettes to health risks the same as "light" cigarettes.

111. Based on the Market-Based method, I estimate that the market value of the change in perceived health risks from greater than regular cigarettes to the same as "light" cigarettes is substantially more than 50 percent of the price per pack.

112. Based on the methodologies described in this report, derived from both consumer Willingness-to-Pay and Market-Based simulations, I conclude, on the basis of the best available information and methodologies, that individual consumers and the market value of the change in perceived health risks from that of a regular cigarette to that of a "light" cigarette is between 39.8 percent and 47.3 percent of the price per pack.

113. Based on the above analysis of the partworths, at most 8/10^{ths} of 1 percent of the respondents use a non-compensatory lexicographic decision rule for taste, health risks, pack type, and price. For all other respondents and for the features of taste, health risks, and price, high values on some features can compensate for low values on other features.

114. I followed accepted scientific methodology to design, execute, and analyze the Experimental Time Study. In addition, I focused on those respondents who, a priori, I believed to be more able to provide estimates as to the times at which they changed their beliefs about health risks and the times at which they first came to believe that they had been misled by the cigarette companies about health risks.

115. The results of the Experimental Time Study suggest that it is difficult for respondents to estimate the year in which they changed their beliefs about health risks and the year in which they first came to believe that they had been misled by the cigarette companies. I do not have confidence in the estimates of these years as provided by respondents.

Dr. John R. Hauser Date

_____ _____